

# A model to identify urban traffic congestion hotspots in complex networks

Albert Solé-Ribalta,<sup>1</sup> Sergio Gómez,<sup>1</sup> and Alex Arenas<sup>1</sup>

<sup>1</sup>*Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, 43007 Tarragona, Spain*

Traffic congestion is one of the most notable problems arising in worldwide urban areas, importantly compromising human mobility and air quality. Current technologies to sense real-time data about cities, and its open distribution for analysis, allow the advent of new approaches for improvement and control. Here, we propose an idealized model, the Microscopic Congestion Model, based on the critical phenomena arising in complex networks, that allows to analytically predict congestion hotspots in urban environments. Results on real cities' road networks, considering, in some experiments, real-traffic data, show that the proposed model is capable of identifying susceptible junctions that might become hotspots if mobility demand increases.

## INTRODUCTION

Urban life is characterized by a huge mobility, mainly motorized. Amidst the complex urban management problems there is a prevalent one: traffic congestion. Several approaches exist to efficiently design road networks[1] and routing strategies[2], however, the establishment of collective actions, given the complex behavior of drivers, to prevent or ameliorate urban traffic congestion is still at its dawn. Usually, congestion is not homogeneously distributed around all city area but there are salient locations where congestion is settled. We call this locations congestion hotspots. These hotspots usually correspond to junctions and are problematic for the efficiency of the network as well as for the health of pedestrians and drivers. It has been shown[3] that drivers in-queue in a traffic jam are the most affected individuals to car exhaust pollution inhalation. In addition, these hotspots are usually located in the city center, magnifying the problem[4]. Assuming that congestion is an inevitable consequence of urban motorized areas, the challenge is to develop strategies towards a sustainable congestion regime at which delays and pollution are under control. The first step to confront congestion is the modelling and understating of the congestion phenomena.

The modeling of traffic flows it is prevalent hot topic since the late 70's when the Gipps' model appear [5]. The Gipps' model and other car-following models [6, 7] have evidenced the necessity of modeling traffic flows to improve road network efficiency and also have shown how congestion severely affects the traffic flows. Since ten years ago also the complex networks' community has proposed stylized models to analyze the problem of traffic congestion in networks and design optimal topologies to avoid it[8–21]. The focus of attention of the previous works was the onset of congestion, which corresponds to a critical point in a phase transition, and how it depends on the topology of the network and the routing strategies used. However, the proper analysis of the system after congestion has remained analytically slippery. It is known that when a transportation network reaches congestion, the system becomes highly non-linear, large fluctuation exists and the travel time and the amount of vehicles queued in a junction diverge[16]. This phenomenon is equivalent to a phase transition in physics, and its modeling is challenging[22–24].

Here, we propose an idealized model to predict the behaviour of transportation networks after the onset of congestion. The presented model is analytically tractable and can be iteratively solved up to convergence. To the best of our knowledge, this is the first analytical model that is able to give predictions beyond the onset of congestion. We present the model in terms of road transportation networks but it could also be applied to analyze other types of transportation networks such as: computer networks, business organizations or social networks.

## RESULTS

To identify congestion hotspots in urban environments we propose a model based on the theory of critical (congestion) phenomena on complex networks. The model, that we call *Microscopic Congestion Model* (MCM), is a mechanistic model (yet simple) and analytically tractable. It is based on assuming that the growth of vehicles observed in each congested node of the networks is constant. This usually happens in real transportation networks at the stationary state. The assumption allows us to describe, with a set of balance equations (one for each node), the increment of vehicles in the junction queues' and the number of vehicles arriving or traversing each junction from neighboring junctions. Mathematically, the increment of the vehicles per unit time at every junction  $i$  of the city,  $\Delta q_i$ , satisfies the following balance equation:

$$\Delta q_i = g_i + \sigma_i - d_i, \quad (1)$$

where  $g_i$  is the average number of vehicles entering junction  $i$  from the area surrounding  $i$ ,  $\sigma_i$  is the average number of vehicles that arrive to junction  $i$  from the adjacent links of that junction, and  $d_i \in [0, \tau_i]$  corresponds to the average of vehicles that actually finish in junction  $i$  or traverse towards other junctions. Note that the value of  $d_i$  is upper-bounded by the maximum amount of vehicles  $\tau_i$  that can traverse junction  $i$  in a time step. This simulates the physical constraints of the road network. A graphical explanation of the variables of the model is shown in Fig. 1.

The system of Eqs. 1 defined for every node  $i$ , is coupled through the incoming flux variables  $\sigma_i$ , that can be expressed

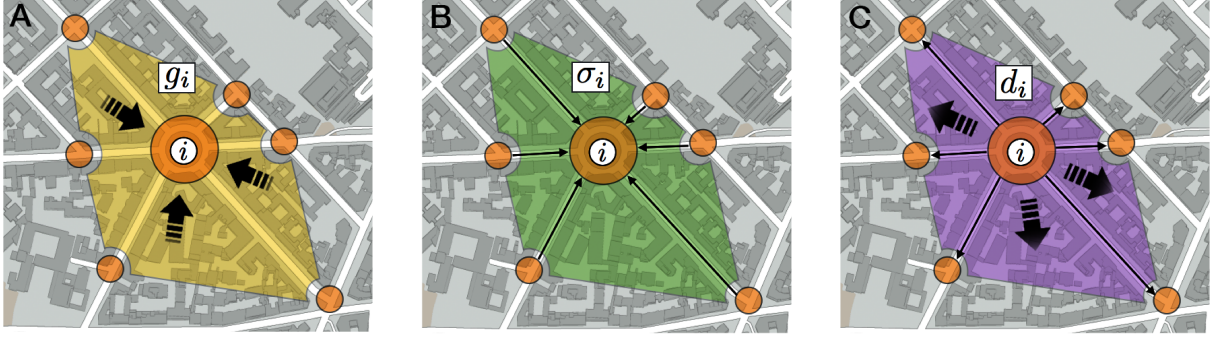


FIG. 1. Illustration of the variables of the MCM model. (A) Vehicles entering junction  $i$  from the area surrounding  $i$ . (B) Vehicles entering junction  $i$  from its neighboring junctions. (C) Vehicles leaving junction  $i$ , either to go to other neighboring junctions or to finishing the trip in its surrounding area.

as

$$\sigma_i = \sum_{j=1}^S P_{ji} p_j d_j, \quad (2)$$

where  $P_{ji}$  accounts for the routing strategy of the vehicles (probability of going from  $j$  to  $i$ ),  $p_j$  stands for the probability of traversing junction  $j$  but not finishing at  $j$  and  $S$  is the number of nodes in the network (see *Materials and Methods* for a detailed description of the MCM).

For each junction  $i$ , the onset of congestion is determined by  $d_i = \tau_i$ , meaning that the junction is behaving at its maximum capability of processing vehicles. Thus, for any flux generation ( $g_i$ ), routing strategy ( $P_{ij}$ ) and origin-destination probability distribution, Eqs. 1 can be solved using an iterative approach (see *Materials and Methods*) to predict the increase of vehicles per unit time at each junction of the network ( $\Delta q_i$ ). The only hypothesis we use is that the system dynamics has reached a stationary state in which the growth of the queues is constant. It is worth commenting here that the MCM model considers a fixed average of new vehicles entering the system  $g_i$ . However,  $g_i$  certainly changes during day time, with increasing values in rush hours and lower values during off-peak periods. MCM can easily consider evolving values of  $g_i$  provided the time scale to reach the stationary state in the MCM (which is usually of the order of minutes in real traffic systems) is shorter than the rate of change in the evolution of  $g_i$  (which is usually of the order of hours for the daily peaks).

#### Validation on synthetic networks

To validate MCM we conducted experiments on several synthetic networks and with two different routing strategies: local search strategy and shortest path strategy. In both routing strategies we assume, for simplicity, that all vehicles randomly choose the starting and ending junctions of their journey uniformly within all junctions of the network. Thus, each

junction generates new vehicles with the same rate  $g_i = \rho$ . For shortest path strategy, vehicles follow a randomly selected shortest path towards the destination. Without loss of generality we fix  $\tau = 1$  and analyze the performance of MCM for different values of  $\rho$ .

Figure 2 shows the accuracy on predicting the values of the order parameter  $\eta = \frac{\sum \Delta q_i}{\rho S}$  and  $d_i$  for shortest paths routing strategy. As in refs. [25, 26], this order parameter  $\eta$  corresponds to the ratio between in-transit and generated vehicles. All experiments show that the MCM achieves high accuracy in predicting the macroscopic and microscopic variables of the stylised transportation dynamics.

#### Application to real scenarios

INRIX Traffic Scorecard (<http://www.inrix.com/>) reports the rankings of the most congested countries worldwide in 2014. US, Canada and most of the European countries are in the top 15, with averages that range from 14 to 50 hours per year wasted in congestion, with their corresponding economical and environmental negative consequences. To demonstrate that the MCM model can be applied to real scenarios to obtain real predictions, in the following we apply the MCM model to the ninth most congested cities according to the INRIX Traffic Scorecard (see Table I).

We first focus on the city of Milan, the city with largest INRIX value. To evaluate the outcome of the MCM model, we first gather data about the road network topology using Open Street Map (OSM). OSM data represents each road (or way) with an ordered list of nodes which can either be road junctions or simply changes of the direction of the road. We have obtained the required abstraction of the road network building a simplified version of the OSM data which only accounts for road junctions (nodes). Then, for each pair of adjacent junctions we have queried the real travel distance (i.e. following the road path) using the API provided by Google Maps. The

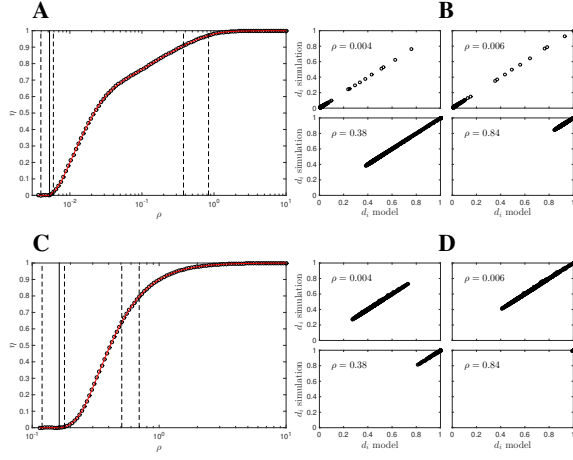


FIG. 2. Validation of the Microscopic Congestion Model with Barabási-Albert (**A**, **B**) and Erdős-Rényi (**C**, **D**) networks of 1000 nodes and shortest path routing strategy. In the construction procedure of the Barabási-Albert networks each new node is connected to 1 existing node in the network. The Erdős-Rényi networks have an average degree of 50. In **A** and **C**, accuracy in predicting the order parameter  $\eta$ . In **B** and **D**, correlation between predicted and simulated values of  $d_i$ . Vertical solid lines on plot **A** and **C** show the predicted critical generation rate  $\rho_c$  (see *Materials and Methods*). Vertical dashed lines show the  $\rho$  values where  $d$  is evaluated on the panels **B** and **D**.

resulting network corresponds to a spatial weighed directed network[27] where the driving directions are represented and the weight of each link indicates the expected traveling time between two adjacent junctions.

Second, we build up the dynamics of the model analyzing real traffic data provided by Telecom Italia for their Big Data Challenge. The data provides, for every car entering the cordoning pricing zone in Milan during November and December 2013, an encoding of the car's plate number, time and gate of entrance (a total of 9183475 records). This allows us to obtain the (hourly) average incoming and outgoing traffic flow, for each gate of the cordoning taxed area.

Given the previous topology and traffic information, we generated traffic compatible with the observations, and evaluated the outcome of the MCM model. Specifically, the simulated dynamics is as follows: for each vehicle entering the Area-C we fix a randomly selected location as destination and use the shortest path route towards it. After the vehicle has arrived to its destination, it randomly chooses an exit door and travels to it also using the shortest path route. This is similar to the well-known *Home-to-Work* travel pattern (see details in *Materials and Methods*). Figures 3 and 4 show the obtained results. Figures 3**B** displays the predicted congestion hotspots on a map of Milan, panel **A** of the same figure shows a real traffic situation obtained with google maps. We see that the predicted congestion hotspots are located in the circular roads of Milan as well as on the arterial roads of the city; this agrees with the real traffic situation shown in panel

TABLE I. Comparison between the INRIX (12 Months) traffic index and the number of hotspots estimated by the proposed model for the most congested cities of the world.

City	INRIX <sup>a</sup>	hotspots	nodes	links
Milano	36.2	108	6924	14315
London	32.4	93	6378	14662
Los Angeles	32.2	57	6799	19368
Brussels	30.5	50	6645	15624
Antwerpen	28.6	44	6530	15252
San Francisco	27.9	45	8854	25530
Stuttgart	21.9	34	8330	19946
Nottingham	21.6	28	7337	16723
Karlsruhe	21.3	19	4257	10379

<sup>a</sup> The INRIX index is the percentage increase in the average travel time of a commute above free-flow conditions during peak hours, e.g. an INRIX index of 30 indicates a 40-minute free-flow trip will take 52 minutes. Each city has been mapped to a graph with the indicated numbers of nodes and links.

**A.** Figures 4**A** shows the distribution of the mean increments each junction has to deal with. This might be a good indicator to decide about future planning actions to improve city mobility. However, differently from what is described in [11], the improvement of the throughput of a single junction might not be enough to improve city mobility since this might end up with the collapse of neighbouring junctions (their incoming rate  $\sigma_i$  will increase). This situation is similar to the Braess' paradox [28]. Figures 4**B** shows the mean increment of vehicles (in vehicles per minute) for each hour of the week-day. The figure clearly shows the morning and evening rush hours as well as the lunch time.

For the other top nine congested cities, we do not have previous traffic information, neither about the real flux of vehicles nor about the vehicle source and destination distributions (to obtain a fair comparison between all the analysed cities we have not consider the Telecom traffic data for Milan here). Thus, for each city, we consider homogeneously distributed source and destination locations and the required road traffic to obtain an order parameter  $\eta$  compatible with the congestion effects recorded by INRIX sensing of real traffic. By relating the INRIX value and  $\eta$ , we are assuming that there exists a relation between the fraction of global congestion and the fraction of extra time wasted reported by INRIX. The obtained results are summarized in Table I, which shows that the amount of hotspots is correlated with the INRIX value. This shows evidences that the percentage increase in the average travel time to commute between to city locations is related to the number of congestion hotspots and with the excess of vehicles within the city.

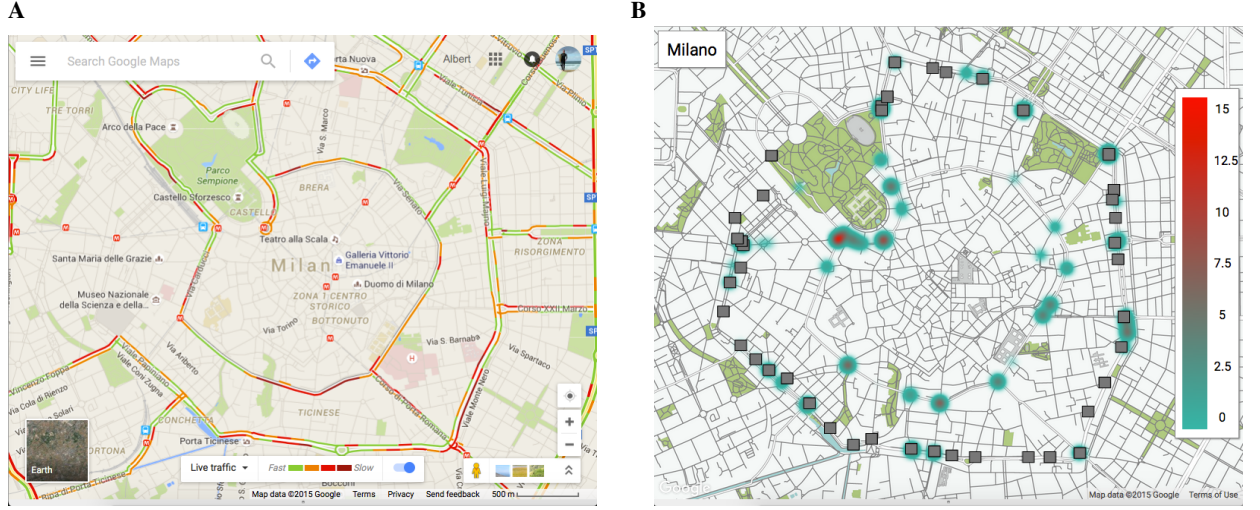


FIG. 3. Congestion hotspot analysis of the city of Milan. Panel **A** shows the typical situation around 9 a.m. for a week day. The image and the data has been obtained with Google Maps. Google maps displays traffic information considering historical data and real-time car velocity reported by smartphones [29]. Panel **B** shows the prediction of the MCM model considering the real road topology obtained using Open Street Map and real traffic data provided by Telecom Italia for their Big Data Challenge. For all congestion hotspots the model has predicted, we shown its mean increment of the queue size,  $\langle \Delta q_i \rangle$ .

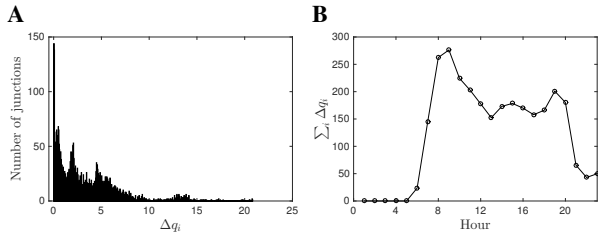


FIG. 4. Statistics of Milan congestion hotspots. Panel **A** shows the distribution of the vehicle increments ( $\Delta q_i$ ) of each congestion hotspot predicted by the MCM. The plot aggregates the MCM predicted congestion considering the average traffic for every week day and every hour of the day. Panel **B** shows the average city congestion in terms of mean increment of vehicle in transit for every hour of the day during a week day.

## DISCUSSION

The previous results show that the MCM (Microscopic Congestion Model) can be used to predict the local congestion before and beyond the onset of congestion of a transportation network. Up to the knowledge of the authors, this is the first analytical model that is able to give predictions beyond the onset of congestion where the system is highly non-linear, large fluctuation exists and the amount of vehicles on transit diverge with respect to time. Our model is based on assuming that the growth of vehicles observed in each congested node of the networks is constant, which allowed us to derive a set of balance equations that can accurately predict microscopic, mesoscopic and microscopic variables of the transportation network.

Traffic congestion is a common and open problem whose

negative impacts range from wasted time and unpredictable travel delays to a waste of energy and an uncontrolled increase of air pollution. A first step towards the understanding and fight of congestion and its related consequences is the analytical modelling of the congestion phenomena. Here, we have shown that the MCM model is detailed enough to give real predictions considering real traffic data and topology. These results pave the way to a new generation of stylized physical models of traffic on networks in the congestion regime, that could be very valuable to assess and test new traffic policies on urban areas in a computer simulated scenario.

## MATERIALS AND METHODS

### Microscopic Congestion Model

Let node  $i$  denote a road junction, edge  $a_{ij}$  the road segment between junctions  $i$  and  $j$ ,  $N_i^{\text{in}}$  and  $N_i^{\text{out}}$  the sets of incoming and outgoing neighbouring junctions of junction  $i$  respectively, and  $S$  the number of junctions in the road network of the city. Incoming vehicles to node  $i$  at each time step can be of two types: those coming from other junctions  $N_i^{\text{in}}$  and those that start its trip with node  $i$  as its first crossed junction. We consider this second type of incoming vehicles as generated in node  $i$ . Our Microscopic Congestion Model (MCM) describes the increment of the vehicles per unit time at every junction  $i$  of the city,  $\Delta q_i$ , as:

$$\Delta q_i = g_i + \sigma_i - d_i, \quad (3)$$

where  $g_i$  is the average number of vehicles generated in node  $i$ ,  $\sigma_i$  is the average number of vehicles that arrive to node  $i$



from  $N_i^{\text{in}}$  junctions, and  $d_i \in [0, \tau_i]$  corresponds to the average number of vehicles that actually finish in it, or traverse this junction towards neighboring nodes in  $N_i^{\text{out}}$ . Parameter  $\tau_i$  represents the maximum routing rate of junction  $i$ . As described in the main text, we decompose the incoming flux of vehicles  $\sigma_i$  to node  $i$  as

$$\sigma_i = \sum_{j \in N_i^{\text{in}}} P_{ji} p_j d_j, \quad (4)$$

where  $p_i$  is the probability that a vehicle waiting in node  $i$  has not arrived to its destination (i.e., it is going to visit at least another junction in the next step) and  $P_{ji}$  is the probability that a vehicle crossing node  $j$  goes to node  $i \in N_j^{\text{out}}$  in its next step.

Since vehicles just generated in a certain node are not affected by the congestion in the rest of the network, we separate their contributions in the computation of probabilities  $p$  and  $P$ . Thus, we decompose  $p_i$  as

$$p_i = p_i^{\text{gen}} p_i^{\text{loc}} + (1 - p_i^{\text{gen}}) p_i^{\text{ext}}, \quad (5)$$

where the first term accounts for vehicles generated in node  $i$  ( $p_i^{\text{gen}}$ ) whose destination is not  $i$  ( $p_i^{\text{loc}}$ ) and the second term accounts for vehicles not generated in  $i$  whose destination is not  $i$  ( $p_i^{\text{ext}}$ ). Supposing trips consist in traveling through two or more junctions we have that  $p_i^{\text{loc}} = 1$ . Probability  $p_i^{\text{gen}}$  is equal to the fraction of vehicles generated in  $i$  with respect to the total amount of incoming vehicles:

$$p_i^{\text{gen}} = \frac{g_i}{g_i + \sigma_i}. \quad (6)$$

Considering the distribution of origins, destinations, the routing strategy and the congestion in the network, probability  $p_i^{\text{ext}}$  can be expressed in terms of the effective node betweenness  $\tilde{B}_i$  and the effective vehicle arrivals  $\tilde{e}_i$  (the amount of vehicles with destination node  $i$  that arrive to node  $i$  at each time step):

$$p_i^{\text{ext}} = \frac{\tilde{B}_i}{\tilde{B}_i + \tilde{e}_i}. \quad (7)$$

The effective betweenness  $\tilde{B}_i$  of a node  $i$  accounts for the expected amount of vehicles each node  $i$  receives per unit time considering the routing algorithm and the overall congestion of the network. See *Materials and Methods* subsection *Effective betweenness in congested transportation networks* for an extended description and computation of the effective node betweenness  $\tilde{B}_i$  and the effective vehicle arrivals  $\tilde{e}_i$ .

In the same spirit, we decompose the probability  $P_{ji}$  that a vehicle waiting in node  $j$  goes to node  $i$  as:

$$P_{ji} = p_j^{\text{gen}} P_{ji}^{\text{loc}} + (1 - p_j^{\text{gen}}) P_{ji}^{\text{ext}}. \quad (8)$$

The first term corresponds to the routed vehicles generated in node  $j$  ( $p_j^{\text{gen}}$ ) that go to node  $i$  ( $P_{ji}^{\text{loc}}$ ) and the second term to the routed vehicles not generated in  $j$  that go to node  $i$  ( $P_{ji}^{\text{ext}}$ ). Similarly as before,  $p_j^{\text{gen}}$  can be expressed as the rate

between the vehicles generated in  $j$  and the total amount of routed vehicles:

$$p_j^{\text{gen}} = \frac{g_j}{g_j + \sigma_j p_j^{\text{ext}}}, \quad (9)$$

and,  $P_{ji}^{\text{loc}}$  and  $P_{ji}^{\text{ext}}$  can be computed in terms of the normalized effective edge betweenness of the network:

$$P_{ji}^{\text{loc}} = \frac{\tilde{E}_{ji}^{\text{loc}}}{S}, \quad (10)$$

$$P_{ji}^{\text{ext}} = \frac{\tilde{E}_{ji}^{\text{ext}}}{S}, \quad (11)$$

where the computation of  $\tilde{E}_{ji}^{\text{loc}}$  only considers paths that start on node  $j$  and  $\tilde{E}_{ji}^{\text{ext}}$  only considers paths that do not start on node  $j$ . Equivalently to the effective node betweenness  $\tilde{B}_i$ , computation of  $\tilde{E}_{ji}^{\text{loc}}$  and  $\tilde{E}_{ji}^{\text{ext}}$  consider, if required, all congested junctions in the network, as described in a later section, as well as the distribution of the vehicle sources and destinations. Note that the sum of  $\tilde{E}_{ji}^{\text{loc}}$  and  $\tilde{E}_{ji}^{\text{ext}}$  corresponds to the classical edge betweenness. Moreover,  $P_{ji}$  is an exact expression before and after the onset of congestion.

Eventually, the MCM is composed by a set of  $S$  equations ( $\Delta q_i = g_i + \sigma_i - d_i$ ), one for each junction, and, in principle, a set of  $2S$  unknowns,  $\Delta q_i$  and  $d_i$  for each junction. To see that the system is indeed determined we need to note that for congested junctions  $\Delta q_i > 0$  and, thus, after the transient state,  $d_i = \tau_i$ . For the non-congested junctions we have that  $\Delta q_i = 0$  and consequently  $d_i = g_i + \sigma_i$ . In conclusion, for any node  $i$ , either  $d_i = \tau_i$  or  $d_i = g_i + \sigma_i$  which reduces the amount of unknowns to  $S$ .

To solve the model given a fixed generation rate  $g_i$ , we start by considering that no junction is congested and we solve the set of equations Eqs. 3–11 by iteration. It is possible that some nodes exceed their maximum routing rate. If this is the case, we set the node with maximum  $d_i$  as congested and we solve the system again. This process is repeated until no new junction exceeds its maximum routing rate.

### Onset of congestion using the Microscopic Congestion Model

Most of the works that consider static routing strategies assume that the generation rate of vehicles is the same for all nodes,  $g_i = \rho$ . In that case, it is possible to compute the critical generation rate  $\rho_c$  such that for any generation rate  $\rho > \rho_c$  the network will not be able to route or absorb all the traffic [25, 26, 30–33]. After this point is reached, the amount of vehicles  $Q(t)$  in the network will grow proportionally with time,  $Q(t) \propto t$ , since some of the vehicles get stacked in the queues of the nodes. This transition to the congested state is

characterized using the following order parameter:

$$\eta(\rho) = \lim_{t \rightarrow \infty} \frac{\langle \Delta Q \rangle}{\rho S}, \quad (12)$$

where  $\langle \Delta Q \rangle$  represents the average increment of vehicles per unit of time in the stationary state. Basically, the order parameter measures the ratio between in-transit and generated vehicles.

In the non-congested phase, the amount of incoming and outgoing vehicles for each node can be computed in terms of the node's algorithmic betweenness  $B_i$ , see ref. [26]. In particular,

$$\sigma_i = \rho \left( \frac{B_i}{S-1} + 1 \right), \quad (13)$$

where the second term inside the parentheses accounts for the fact that, in our model, vehicles are also queued at the destination node, unlike in ref. [26]. When no junction is congested we have that  $\Delta q_i = 0$  for all nodes and consequently

$$d_i = \rho + \sigma_i = \rho \left( \frac{B_i}{S-1} + 2 \right). \quad (14)$$

A node  $i$  becomes congested when it is required to process more vehicles than its maximum processing rate,  $d_i > \tau$ . Thus, the critical generation rate at which the first node, and so the system, reaches congestion is:

$$\rho_c = \min_i \frac{\tau(S-1)}{B_i + 2(S-1)}. \quad (15)$$

This is one of the most important analytical results on transportation networks with static routing strategies. In the following, we show that we can recover Eq. (13) before the onset of congestion using our MCM approach. After substitution of the expression of the probabilities in Eq. 4:

$$\sigma_i = \sum_j \frac{\rho(B_j + S - 1)P_{ji}^{\text{loc}} + \sigma_j B_j P_{ji}^{\text{ext}}}{(\rho + \sigma_j)(B_j + S - 1)} d_j, \quad (16)$$

and, given we do not have congestion (i.e.,  $d_j = \rho + \sigma_j$ ), it simplifies to

$$\sigma_i = \sum_j \frac{\rho(B_j + S - 1)P_{ji}^{\text{loc}} + \sigma_j B_j P_{ji}^{\text{ext}}}{B_j + S - 1}. \quad (17)$$

Equation (17) in matrix form becomes

$$(I - M)\sigma = \rho\pi, \quad (18)$$

where

$$M_{ij} = \frac{B_j P_{ji}^{\text{ext}}}{B_j + S - 1}, \quad (19)$$

$$\pi_i = \sum_j P_{ji}^{\text{loc}}, \quad (20)$$

and then

$$\sigma = \rho(I - M)^{-1}\pi. \quad (21)$$

This expression can be shown to be equivalent to Eq. (13) by using the following relationship between node and edge betweenness:

$$B_i + (S - 1) = \sum_j (B_j P_{ji}^{\text{ext}} + (S - 1)P_{ji}^{\text{loc}}). \quad (22)$$

The right hand side corresponds to the accumulated fractions of paths that pass through the neighbors of node  $i$  and then go to  $i$ . Each neighbor contributes with two terms, the paths that go through  $j$  coming from other nodes, and the paths that start in  $j$ .

### Effective betweenness in congested transportation networks

The effective betweenness  $\tilde{B}_i$  of a node  $i$ , as defined in ref. [26], accounts for the expected amount of vehicles each node  $i$  receives per unit time. When the network is not congested and the vehicle generation rate  $g_i$  is equal for all nodes,  $g_i = \rho$ , the number of vehicles each node receives can be obtained using Eq. 13. However, if the network is congested, the traffic dynamics becomes highly non-linear and the value of  $\sigma_i$  computed in Eq. 13 becomes a poor approximation.

Suppose we focus on a particular congested node  $j^*$  of the network. For  $j^*$ , being congested means that it is receiving more vehicles than the ones it can process and route. In particular, from the  $\sigma_{j^*} + g_{j^*}$  vehicles that arrive to the node, only  $\tau_{j^*}$  can be processed at each time step.

Therefore, the contribution to the effective betweenness  $\tilde{B}_i$  of the paths from a source/destination pair,  $(s, t)$ , that traverse the congested node  $j^*$  before reaching  $i$ , must be rescaled by the fraction of processable vehicles:

$$s_{j^*} = \frac{\tau_{j^*}}{\sigma_{j^*} + g_{j^*}}. \quad (23)$$

When a path traverses multiple congested nodes  $j^*, k^*, \dots$ , the remaining fraction of paths that will reach the target node will be the result of the application of the multiple re-scalings  $s_{x^*}$ .

The computation of  $s_{j^*}$  is not straightforward. In general,  $\sigma_i$  is not known after the onset of congestion and depends on the effective betweenness that requires, at the same time, to know the  $s_{j^*}$  fraction for all congested nodes. Thus, an iterative calculation is needed to fit all the parameters at the same time as we do in our Microscopic Congestion Model.

The effective arrivals  $\tilde{e}_i$  account for the amount of vehicles with destination node  $i$  that arrive to node  $i$  at each time step. This value in the non-congested phase can be obtained, considering homogeneous source and destination nodes, as

$$e_i = \rho(S - 1). \quad (24)$$

However, congestion affects the variable  $e_i$  as well, and needs to be corrected accordingly using the same procedure presented above.

## Traffic Dynamics

To simulate the traffic dynamics of the road network, we assign a first-in-first-out queue to each junction that simulates the blocking time of vehicles before they are allowed to cross it and continue their trip. We suppose these queues have infinite capacity and a maximum processing rate that simulates the physical constraints of the junction. Vehicles origins and destinations may follow any desired distribution. In this work, we have considered two distributions: a random uniform distribution for the synthetic experiments, and one obtained considering the ingoing and outgoing flux of vehicles of the city of Milan. At each time step (of 1 minute duration) vehicles are generated and arrive to their first junction. During the following time steps, vehicles navigate towards their destination following any routing strategy. Here, we have used two different routing strategies: shortest-path and random local search.

For the particular case of simulating traffic in the city of Milan we assume a traffic dynamics similar to the “Home-to-Work” travel pattern where vehicles arrive from the outskirts of the city, go to the city center and then return to the outskirts. Specifically, in our simulation, traffic is generated in the peripheral junctions of the network, goes to a randomly selected junction within the city and then returns back to a randomly selected peripheral junction. We do not consider trips with origin and destination inside the city center since public transportation systems (e.g., train or subway) usually constitute a better alternative than private vehicles for those trips.

The maximum crossing rate of each junction  $\tau_i$  accounts, among others, for the existence of traffic lights governing the junction, the width of the street as well as its traffic. We have not been able to get this information for the studied cities, and consequently we cannot set to each junction its precise value. Instead, without loss of generality and for the sake of simplicity, we set to all junctions the same maximum crossing rate,  $\tau_i = 15$  (an estimation of the average of their real values).

## ACKNOWLEDGEMENTS

This work has been supported by Ministerio de Economía y Competitividad (Grant FIS2015-71582-C2-1) and European Commission FET-Proactive Projects MULTIPLEX (Grant 317532). A.A. also acknowledges partial financial support from the ICREA Academia and the James S. McDonnell Foundation.

- 
- [1] H. Yang, M. G. H. Bell, Models and algorithms for road network design: a review and some new developments. *Transport Reviews* **18**, 257-278 (1998).
  - [2] H. Bast, S. Funke, P. Sanders, D. Schultes, Fast routing in road networks with transit nodes. *Science* **316**, 566–566 (2007).
  - [3] G. Petersson, *Traffic and children's health* (NHV-Report 1987:2, 1987), chap. Exposure to traffic exhaust, pp. 117–126.
  - [4] G. Raducan, S. Stefan, Characterization of traffic-generated pollutants in bucharest. *Atmósfera* **22**, 99–110 (2009).
  - [5] R. E. Wilson, An analysis of gipp's car-following model of highway traffic. *IMA journal of applied mathematics* **66**, 509–537 (2001).
  - [6] M. Treiber, A. Hennecke, D. Helbing, Congested traffic states in empirical observations and microscopic simulations. *Physical Review E* **62**, 1805 (2000).
  - [7] G. F. Newell, A simplified car-following theory: a lower order model. *Transportation Research Part B: Methodological* **36**, 195–205 (2002).
  - [8] R. Guimerà, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, A. Arenas, Optimal network topologies for local search with congestion. *Phys. Rev. Lett.* **89**, 248701 (2002).
  - [9] B. Tadić, S. Thurner, G. Rodgers, Traffic on complex networks: Towards understanding global statistical properties from microscopic density fluctuations. *Physical Review E* **69**, 036102 (2004).
  - [10] L. Donetti, P. I. Hurtado, M. A. Muñoz, Entangled networks, synchronization, and optimal network topology. *Phys. Rev. Lett.* **95**, 188701 (2005).
  - [11] L. Zhao, Y.-C. Lai, K. Park, N. Ye, Onset of traffic congestion in complex networks. *Phys. Rev. E* **71**, 026125 (2005).
  - [12] D. J. Ashton, T. C. Jarrett, N. F. Johnson, Effect of congestion costs on shortest paths through complex networks. *Physical review letters* **94**, 058701 (2005).
  - [13] B. K. Singh, N. Gupte, Congestion and decongestion in a communication network. *Physical Review E* **71**, 055103 (2005).
  - [14] B. Danila, Y. Yu, J. A. Marsh, K. E. Bassler, Optimal transport on complex networks. *Phys. Rev. E* **74**, 046106 (2006).
  - [15] M. Barthélemy, A. Flammini, Optimal traffic networks. *Journal of Statistical Mechanics: Theory and Experiment* **2006**, L07002 (2006).
  - [16] S. N. Dorogovtsev, A. V. Goltsev, J. F. Mendes, Critical phenomena in complex networks. *Reviews of Modern Physics* **80**, 1275 (2008).
  - [17] K. Kim, B. Kahng, D. Kim, Jamming transition in traffic flow under the priority queueing protocol. *EPL (Europhysics Letters)* **86**, 58002 (2009).
  - [18] G. Li, *et al.*, Towards design principles for optimal transport networks. *Phys. Rev. Lett.* **104**, 018701 (2010).
  - [19] J. J. Ramasco, S. Marta, E. López, S. Boettcher, Optimization of transport protocols with path-length constraints in complex networks. *Physical Review E* **82**, 036119 (2010).
  - [20] S. Scellato, L. Fortuna, M. Frasca, J. Gómez-Gardeñes, V. Latora, Traffic optimization in transport networks based on local routing. *The European Physical Journal B* **73**, 303–308 (2010).
  - [21] J. Gleeson, S. Melnik, J. A. Ward, M. Porter, P. Mucha, Accuracy of mean-field theory for dynamics on real-world networks. *Physical Review E* **85**, 026106 (2012).
  - [22] D. De Martino, L. Dall'Asta, G. Bianconi, M. Marsili, A minimal model for congestion phenomena on complex networks. *Journal of Statistical Mechanics: Theory and Experiment* **2009**, P08023 (2009).
  - [23] P. Echenique, J. Gómez-Gardeñes, Y. Moreno, Dynamics of jamming transitions in complex networks. *EPL (Europhysics Letters)* **71**, 325 (2005).
  - [24] L. Lü, L. Pan, T. Zhou, Y.-C. Zhang, E. Stanley, Toward link predictability of complex networks. *Proceedings of the National Academy of Sciences* **112**, 2325–2330 (2015).
  - [25] A. Arenas, *et al.*, Optimal information transmission in organizations: search and congestion. *Review of Economic Design* **14**, 75–93 (2010).

- [26] R. Guimerà, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, A. Arenas, Optimal network topologies for local search with congestion. *Physical review letters* **89**, 248701 (2002).
- [27] M. Barthélemy, Spatial networks. *Physics Reports* **499**, 1–101 (2011).
- [28] H. Youn, M. T. Gastner, H. Jeong, Price of anarchy in transportation networks: efficiency and optimality control. *Physical review letters* **101**, 128701 (2008).
- [29] D. Barth, Product Manager for Google Maps, <https://googleblog.blogspot.in/2009/08/bright-side-of-sitting-in-traffic.html> (2009). [Online; accessed 17-December-2015].
- [30] A. Arenas, A. Díaz-Guilera, R. Guimera, Communication in networks with hierarchical branching. *Physical Review Letters* **86**, 3196 (2001).
- [31] D. De Martino, L. Dall’Asta, G. Bianconi, M. Marsili, Congestion phenomena on complex networks. *Physical Review E* **79** (2009).
- [32] S. Chen, W. Huang, C. Cattani, G. Altieri, Traffic dynamics on complex networks: a survey. *Mathematical Problems in Engineering* **2012** (2011).
- [33] A. Arenas, A. Cabrales, A. Díaz-Guilera, R. Guimerà, F. Vega-Redondo, *Statistical Mechanics of Complex Networks* (Springer, 2003), pp. 175–194.